Monitoring income-related health differences between regions in Great Britain

Allanson, Paul

Published in:
Social Science and Medicine

DOI:
10.1016/j.socscimed.2016.12.033

Publication date:
2017

Monitoring income-related health differences between regions in Great Britain: a new measure for ordinal health data

Abstract

The paper proposes a new measure of the extent to which differences in population health status between the regions of a country are systematically related to regional prosperity. The headcount index of income-related health stratification has a straightforward interpretation as the population-weighted mean difference in the probabilities that the healthier of any two randomly chosen individuals will be from the richer rather than the poorer region from which they are drawn. Moreover, it is well-defined even if only ordinal health data are available, being directly applicable to polytomous categorical variables without the need for either dichotomisation or cardinalisation. The new index is used to examine the evolution of income-related health differences between the regions of Great Britain over the period from 1991 to 2008.

Keywords: headcount index; income-related health stratification; regional analysis; ordinal data

JEL classifications: D63, I14, I18

1. Introduction

Improvements in health over recent decades have not generally been matched by reductions in health inequalities between regions, both within and across countries. In Great Britain, for example, there has been a long-running debate (see Taulbut et al., 2013) about why health outcomes have been persistently worse in Scotland than in England and Wales even after controlling for differences in levels of social deprivation – the so-called 'Scottish' or 'Glasgow' effect. The ongoing impact of the financial crisis in 2008 has also renewed concerns about health differences between English regions, leading Public Health England to commission an independent inquiry on health equity for the less prosperous North of the country (see Whitehead, 2014). Across the older (pre-2004) Member States of the European Union, Marmot (2013, Table 3.2) reports that the difference in life expectancy at age 50 between the richest and poorest NUTS 2 regions was about 1 year for both men and women in 2008.

The focus of this paper is on the measurement of the association between regional health outcomes and socioeconomic conditions, which is typically represented graphically by the social gradient in health and summarised numerically by rank-dependent health inequality statistics such as between-region slope inequality and concentration indices. One obvious limitation of existing measurement tools based on the average health and income of each region is that they fail to take into consideration the variation in outcomes that occurs within regions. It is of course possible to calculate both between-region and within-region health inequality measures if suitable data on individual or local district level health and incomes are available. But the sum of the between and within components will not exactly equal the overall level of socioeconomic health inequality if, as will usually be the case, the regional income ranges overlap to any extent, with the additional 'residual' term difficult to interpret in practice (see Jiménez-Rubio et al. 2008). A second more general problem concerns the

question of how to measure health inequalities using ordinal or categorical data, such as survey measures of self-reported health and subjective well-being, without first converting the data into cardinal form by assigning some more or less arbitrary numerical values either to each response category or to the differences between categories (see Allison and Forster, 2004; Lv et al., 2015; Kobus, 2015). As a result, empirical work has focused very largely on differences in cardinal health measures such as life expectancy and prevalence rates (e.g. the percentage of the population with a disability), with evidence on regional disparities in ordinal measures of health and well-being both more limited and equivocal in nature. This paper proposes a new measure of income-related health differences between regions that both takes account of intra-regional variation in health outcomes and is applicable even if only ordinal health data are available.

More specifically, we set out to measure the degree of stratification between the population health distributions of the regions of a country, where this approach contrasts with the conventional focus in health inequalities research on "the evaluation of the inequality in the distribution of health status across individuals in a population" (Allison and Foster, 2004, p.505). The concept of stratification is deeply embedded within sociology, most notably in relation to the analysis of social class, but has only been of relatively recent concern within the economics literature. Thus Yitzhaki and Lerman (1991) in their seminal article quote a definition by the sociologist Lasswell (1965, p.10): "In its general meaning a stratum is a horizontal layer, usually thought of as between, above or below other such layers or strata. Stratification is the process of forming observable layers, or the state of being comprised of layers." Accordingly, we seek to evaluate the degree to which the populations of different regions occupy well-defined strata in the national health distribution, with the socioeconomic dimension taken into account by ranking the regions in terms of economic prosperity rather than population health status. The analysis could be extended to also take into account the

scale of between-region differences in health outcomes if cardinal health data are available, but this lies beyond the scope of the current paper.

Our approach builds on the class of univariate stratification indices proposed by Allanson (2016) for the analysis of income stratification between regions, with the univariate headcount index employed directly to measure 'pure' health stratification between regions. However, the main contribution of the paper is to provide an extension that leads to the specification of a headcount index of income-related health stratification (IRHS), which is related to the univariate index in the same way that the health concentration index is to the health Gini. Thus we first rank regions in terms of economic prosperity and then proceed to measure headcount IRHS between each pair of regions as the difference in the probabilities that a randomly selected individual from the richer region is more rather than less healthy than a randomly chosen individual from the poorer region. This pairwise measure captures the degree to which the two regions form well-defined strata in their combined health distribution, with the headcount index obtained by aggregating over all pairs of regions to yield a national population-weighted average of the pairwise indices. It is demonstrated that the IRHS headcount index satisfies a health status exchange condition that is akin to the Pigou-Dalton principle of transfers in inequality analysis, providing a measure that is equal to twice the between-region generalised health concentration index for binary health status indicators but is also well-defined for polytomous categorical variables.

Our methodology differs from most of the literature on the measurement of health inequality with ordinal data in that it incorporates the socioeconomic dimension, with the seminal paper by Allison and Foster (2004) emphasising the point that their method is designed to evaluate overall inequality in health, without focusing on any particular cause or justification. One major exception is Zheng (2011) who develops a set of welfare dominance and inequality ordering conditions, together with associated summary measures, to compare

socioeconomic inequality in health between pairs of regions given data on individual health and income outcomes. The approach is based on the construction for each region of an income-health matrix that gives the health profile for each of a finite set of income classes, with the various conditions obtained on the assumption that higher income classes have better health prospects than lower classes within each region. A comparison of the US and Canadian socioeconomic health distributions fails to demonstrate either Generalised Lorenz or Lorenz dominance, though both welfare and inequality rankings are obtained through the use of higher order dominance conditions that impose more restrictions on health outcome values (see also Wang and Yu, 2016). The summary indices indicate lower levels of welfare and higher levels of absolute inequality in the US than Canada, though the measures lack simple, everyday interpretations. In contrast our methodology is motivated in this paper by the notion of statistical preference (De Schuymer et al., 2003), which provides a 'graded' alternative to stochastic dominance that yields both a complete ordering of regions and a readily intelligible measure of the differences in population health between them.

The new index is used to examine the evolution of income-related health differences between the regions of Great Britain, making use of a range of self-assessed health measures available in the British Household Panel Survey (BHPS) over the period 1991 to 2008. The next section introduces the headcount IRHS index with the results of the empirical study presented in Section 3. The final section summarises the contribution and offers some suggestions for further applications of the measurement approach.

2. Measurement of headcount income-related health stratification

Consider the population of some country that consists of $R \geq 2$ mutually exclusive and exhaustive administrative regions. The population and population share in region r $(r = 1,....R)$ are given as n_r and $p_r = n_r / N$ respectively, where $N = \sum_r n_r$ is the

national population. Let H_r denote the health variable in region *r*. The probability that a randomly chosen individual from region *r* is (strictly) healthier than a randomly chosen individual from region *s* is given as $\mathrm{P}(H_r > H_s)$.

Our measurement of stratification further requires the prior imposition of some ordering on the regions. Thus, the first step in the construction of a univariate or 'pure' measure of health stratification would be to order regions by population health status. If the health measure is cardinal then this could be done in the manner of Allanson (2016) by ordering regions by mean health with any ties separated on the basis of health distribution ranks such that $\mathrm{P}(H_s > H_r) > \mathrm{P}(H_r > H_s)$ for all relevant pairwise comparisons. If the health measure is ordinal then the secondary criterion may be used on its own, generating a transitive ordering if the probability relationship between the set of regions exhibits mutual rank transitivity (De Baets et al., 2010).

However the main focus of this paper is on the construction of bivariate measures of IRHS so, in the remainder of this section, regions are instead assumed to be ordered by some measure of economic prosperity from the poorest $(r = 1)$ to the richest region $(r = R)$ on the basis of the above rules. For example, regions might be ranked by mean income, with ties separated on the basis of income distribution ranks when necessary, or by the secondary criterion alone if using some area-based index of local neighbourhood income deprivation. Income-related and 'pure' health stratification will be the same if the ordering of regions by economic prosperity and by population health are identical. We note that small changes in individual incomes, for example, may lead to discontinuous changes in IRHS if they lead to changes in the ordering of regions by mean income.

2.1 Measurement of pairwise headcount IRHS

Headcount IRHS between any two regions *r* and *s* depends on the degree to which the populations of the two regions occupy well-defined strata in their combined health distribution as measured by the pairwise identification index I_{rs}:

$$\begin{aligned} I_{rs} &= \text{sgn}(s-r)\left(\text{P}\left(H_s > H_r\right) - \text{P}\left(H_r > H_s\right)\right) \\ &= \text{sgn}(s-r)\left(\left(\text{P}\left(H_s > H_r\right) + 0.5\,\text{P}\left(H_s = H_r\right)\right) - \left(\text{P}\left(H_r > H_s\right) + 0.5\,\text{P}\left(H_s = H_r\right)\right)\right) \\ &= 1 - 2\left(\text{P}\left(H_r > H_s\right) + 0.5\,\text{P}\left(H_s = H_r\right)\right) \end{aligned} \tag{1}$$

$$\text{where} \quad \text{sgn}\left(s-r\right) = \begin{cases} 1 & if\ s-r > 0 \\ 0 & if\ s-r = 0 \\ -1 & s-r < 0 \end{cases} \tag{1a}$$

I_{rs} is thus equal to the signed difference in the probabilities that a randomly chosen individual from region *s* will have better rather than worse health than a randomly chosen individual from region *r*. I_{rs} is symmetric, as the specification of the sign function in (1a) implies $I_{rs} = I_{sr}$, but the index is nevertheless sensitive to the ordering of regions by economic prosperity, providing a 'directional' measure in the sense of Dagum (1997). Thus, in the limiting case of two regions with non-overlapping health distributions, $I_{rs} = 1$ if health in the richer region *s* is better than in the poorer region *r* and $I_{rs} = -1$ if the opposite is true. $I_{rr} = I_{ss} = 0$ by construction.

I_{rs} is defined for both continuous and discrete health distributions, with the second line of (1) making explicit the treatment of ties in the case that $\text{P}\left(H_r = H_s\right) \neq 0$, which will be the norm with self-reported health data from surveys in which individuals are typically asked to choose between a finite number of descriptive categories (e.g. very poor, poor, fair, good, excellent). Thus, importantly, I_{rs} is well defined even if only ordinal health data are available: for example $I_{rs} = (0.36 - 0.16) = 0.2$ if health is given by a binary variable with 40%

and 60% respectively of the region *r* and *s* populations reporting good health. The final line of (1) follows by definition.

The difference in probabilities $\mathrm{P}(H_s > H_r) - \mathrm{P}(H_r > H_s)$ provides the basis for the comparison of the population health of the two regions. Specifically, H_s may be said to be weakly statistically preferred to H_r if this difference is greater than or equal to zero, since the odds are at least evens that a randomly chosen individual from region *s* will have better rather than worse health than one from region *r*. Statistical preference (De Schuymer et al., 2003) has not previously been used to compare population health outcomes, but provides an attractive alternative to stochastic dominance (cf. Zheng, 2011) for this purpose. First, its use appears more natural if data on health outcomes are qualitative in nature since the criterion only entails judgements on whether one health outcome is better, worse or the same as another (see Montes et al., 2015, for further discussion). Second, it will always provide a complete ranking of all pairs of regions whereas stochastic dominance may not, though the resultant ordering need not necessarily be transitive: for example, if $H_r = \{5,2,2\}$, $H_s = \{3,3,3\}$ and $H_t = \{4,4,1\}$, where higher scores imply better health, then $\mathrm{P}(H_s > H_r) = 2/3$, $\mathrm{P}(H_t > H_s) = 2/3$ and $\mathrm{P}(H_r > H_t) = 5/9$. Weak statistical preference provides a generalisation of weak first-degree stochastic dominance, since the latter implies the former, but not vice versa (De Baets and De Mayer, 2007). Finally, statistical preference provides a 'graded' comparison, with I_{rs} offering a readily intelligible measure of the degree to which population health in the richer region is better or worse than in the poorer one.

The interpretation of I_{rs} as an identification or classification index follows from the observation that if individuals from the two regions are randomly matched with each other then I_{rs} will reflect the success with which regional identity can be determined by assuming that the healthier individual within each pair will be from the richer rather than poorer region.

I_{rs} will take its maximum value of one if regional identity can be determined with certainty by this rule, which will only be the case if the least healthy individual in the richer region is more healthy than the most healthy individual in the poorer region: not only will everyone from the richer region be among the healthiest people in the two regions but also all the healthiest people will be from the richer region. Conversely, I_{rs} will equal zero if the health distributions of the two regions are identical such that the pairwise identification rule is entirely uninformative of regional identity: the healthier of any pair is equally likely to be from one region as the other if the two regions are indistinguishable in terms of health outcomes. Finally, I_{rs} will be negative if the healthier of any randomly chosen pair is more likely to be from the poorer than richer region, taking a minimum value of minus one.

2.2. Definition and properties of the IRHS headcount index

The IRHS headcount index *S* is obtained as the population-weighted average of the pairwise identification indices I_{rs}:

$$\begin{aligned} S &= \sum_{r=1}^{R}\sum_{s=1}^{R} p_r p_s I_{rs} = \sum_{r=1}^{R}\sum_{s=1}^{R} p_r p_s \,\mathrm{sgn}(s-r)\left(\mathrm{P}\left(H_s > H_r\right) - \mathrm{P}\left(H_r > H_s\right)\right) \\ &= 2\sum_{r=1}^{R}\sum_{s=r+1}^{R} p_r p_s \left(\mathrm{P}\left(H_s > H_r\right) - \mathrm{P}\left(H_r > H_s\right)\right); \end{aligned} \tag{2}$$

where $p_r p_s$ may be interpreted as the probability that the first of two individuals randomly selected with replacement from the national population will be from region *r* and the second from region *s*, and which therefore sum to one over all possible combinations. *S* thus measures the mean difference in the probabilities that the healthier of two randomly chosen individuals will come from the richer rather than the poorer region in pairwise comparisons.

S will take a value of zero if all pairwise indices I_{rs} are zero, although this does not necessarily imply that all regional health distributions are identical. *S* is strictly increasing in I_{rs}, which provide unique estimates of the contribution of each pair of regions to headcount

IRHS. Thus, the pairwise indices may be meaningfully aggregated, given symmetry, to yield estimates $S_r = p_r \sum_s p_s I_{rs}$ of the contribution of each region to *S*, with the further potential to identify the characteristics or factors that contribute to stratification. *S* is invariant to the permutation of regions and to the replication both of the subpopulations within regions (holding the population shares of the regions constant) and of the regions (holding the subpopulations within each region constant).

S may be interpreted as a headcount or incidence measure that captures the extent to which individuals' positions within the national health distribution are determined by regional prosperity. In particular, S will take its maximum value of $\left(1-\sum_r p_r^2\right)$ if there is both complete separation of the regional populations into discrete layers in the national health distribution and the ordering of the regions by population health is the same as by income. In this case there is perfect stratification in the sense of Laswell (1965), with individuals from any particular region restricted to a single interval or range of ranks in the national health distribution that is exclusively occupied by individuals from their own region. Conversely $S = 0$ if regional prosperity is entirely uninformative as a predictor of relative rank such that $I_{rs} = 0$ for all pairs of regions, though a zero value may also arise in cases in which positive and negative values of the pairwise indices cancel each other out. Negative values of S imply that economic prosperity is negatively correlated at the regional level with population health ranks. Dividing S by $\left(1-\sum_r p_r^2\right)$ yields a normalised index $\hat{S}$ that is the population-weighted mean level of pairwise identification between all mutually distinct regions, with maximum and minimum values of plus and minus one respectively.

S satisfies a health status exchange condition akin to the Pigou-Dalton principle of transfers in health inequality analysis (see Bleichrodt and van Doorslaer, 2006). This condition holds that an exchange in health status (and hence of ranks in the national health

distribution) between an individual from a richer region and an individual in worse or equal health from a poorer region will not lead to an increase in headcount IRHS provided that the exchange does not affect the ordering of regions. Let ΔI_{rs} be the resultant change in pairwise identification for any two regions *r* and *s*, with *s* richer than *r*, then $\Delta I_{rs} \leq 0$ because $\Delta \mathrm{P}\left(H_s > H_r\right) \leq 0$ and $\Delta \mathrm{P}\left(H_r > H_s\right) \geq 0$. Moreover, it is easily shown for any third region *q* that $p_s \Delta I_{sq} + p_r \Delta I_{rq} = 0$, $p_s \Delta I_{qs} + p_r \Delta I_{qr} = 0$ and $p_s \Delta I_{qs} = p_r \Delta I_{rq} \leq 0$ if q is respectively richer than both *s* and *r*, poorer than both *s* and *r*, and poorer than *s* but richer than *r*. In contrast, a simple transfer of health between the two individuals may increase headcount IRHS since, for example, the identification of the richer region might not change as a result while that of the poorer region could increase in relation to even poorer regions. More generally, stratification can readily be distinguished from inequality if health is cardinally measurable since a reduction in within-region health variation holding between-region differences constant will lead to a fall in health inequality according to the principle of health transfers, but a rise in stratification if, as is likely, it reduces the degree of overlap between regional health distributions (see Allanson, 2016, for further discussion).

S is a unit free measure that is continuous in individual health outcomes and invariant to rank-preserving transformations of them. If the health outcome measure is given by a binary indicator variable, taking values of zero and one, then S is equal to twice the conventional between-region generalised concentration index since:

$$\begin{aligned} S &= \sum_{r=1}^{R} \sum_{s=1}^{R} p_r p_s \,\mathrm{sgn}(s-r)\left(\mathrm{P}\left(H_s > H_r\right) - \mathrm{P}\left(H_r > H_s\right)\right) \\ &= \sum_{r=1}^{R} \sum_{s=1}^{R} p_r p_s \,\mathrm{sgn}(s-r)\left(P(H_s = 1)(1 - P(H_r = 1)) - P(H_r = 1)(1 - P(H_s = 1))\right) \\ &= \sum_{r=1}^{R} \sum_{s=1}^{R} p_r p_s \,\mathrm{sgn}(s-r)\left(P(H_s = 1) - P(H_r = 1)\right) \\ &= \sum_{r=1}^{R} \sum_{s=1}^{R} p_r p_s \,\mathrm{sgn}(s-r)\left(\mu_s - \mu_r\right) = 2\mu C_B \end{aligned} \tag{3}$$

where $\mu = \sum_r p_r \mu_r = \sum_r p_r P(H_r = 1)$ may be interpreted as a measure of national mean health and C_B is the between-region health concentration index. But, unlike the between-region (generalised) concentration index, S is also defined for polytomous categorical variables without the need to first impose some essentially arbitrary cardinalisation of the health measure. Moreover, even in the dichotomous case S has a more natural and intuitive interpretation (cf. Wagstaff (2005) and the subsequent exchange of views following Erreygers (2009)).

With only two regions, the reduction in headcount IRHS caused by an incremental improvement in one person's health would be greatest for an individual in the poorer region with health equal to the modal health level in the richer region. With more than two regions, the issue is more complicated as there is a need to consider which region to target as well as to identify which individuals in the targeted region to treat, where this will depend for middle income regions on the net change in identification due to an incremental change in the chosen person's health outcome. Nevertheless it is readily apparent that improving the health of the poorest region, let alone the health of the least healthy individuals in that region, will not necessarily have the most impact on headcount IRHS: indeed S is invariant to changes in the health of individuals in the poorest region whose health is worse, and remains worse, than the most unhealthy individual in any other region. In contrast, it would be sufficient to simply target health improvements at the poorest region to have maximum impact on between-region income-related health inequality.

3. Empirical analysis

The headcount index is used to examine the evolution of income-related health differences between the regions of Great Britain. Our empirical analysis employs data on self-reported health from waves 1 to 18 of the British Household Panel Survey (BHPS; University of

Essex, Institute for Social and Economic Research, 2010)), covering the years 1991 through 2008. Established in 1991, the BHPS was a household panel survey with yearly interviews of all adults in each household covering a range of topics including health, work, education, income, family, and social life. The BHPS was designed to be representative of private households in Great Britain 'at multiple time points corresponding to the waves of data collection' (Buck et al., 2006), with the original panel of approximately 5500 households boosted by the recruitment of new extension samples from Scotland and Wales in 1999. The BHPS was replaced by the successor study, Understanding Society, following wave 18 in 2008.

The study is based on NUTS 1 statistical regions – Wales, Scotland and the nine Government Office Regions in England. Sample weights are used throughout the analysis with these being given by standardised BHPS cross-sectional respondent weights for each wave, where the standardisation takes account of wave-specific regional differences in population structure by sex and five-year age band compared to Great Britain as a whole. Standard errors for all inequality and stratification measures are generated using a bootstrap procedure in which re-sampling is carried out at the cluster (Primary Sampling Unit) rather than individual level within each stratum, reflecting the sample design.

3.1 Regional ordering by income

The ordering of regions by mean income could be determined using income data from the BHPS, but we principally rely instead on information about living standards from Households Below Average Income (HBAI) as this is generally considered to be the most reliable source of evidence on UK household net income and poverty (Office for National Statistics, 2016). Specifically, we use HBAI statistics (Department of Work and Pensions, 2015) on mean equivalised household incomes before housing costs by region as the primary

ordering criterion and, in the few cases of ties, further rank regions on the basis of the distribution of individuals by (national) quintile groups as reported in Regional Trends (Office for National Statistics, various years). The HBAI series is based on data from the Family Resources Survey of the total weekly income of all household members after deductions of income tax and other contributions but before housing costs, with this total being equivalised to take account of the size and composition of the household. Three-year centred moving averages are reported at the regional level as single-year estimates are considered too volatile. Prior to the start of the HBAI series in 1995, we use, as the closest comparable measure calculable from published statistics, three year averages of normal weekly disposable per capita household income by (standard statistical) region based on Family Expenditure Survey data. Rankings for 1991 to 1994 were broadly consistent with those from 1995 onwards.

3.2 General health and health-related quality of life (HRQoL) variables

Respondents have been asked about their general state of health in all waves of the BHPS, but there are differences in the phrasing of this question between waves. In all waves except wave 9, respondents were asked to think about their health over the past 12 months compared to people of their own age and say whether on the whole it had been very poor, poor, fair, good or excellent (BHPS variable: HLSTAT). In contrast, respondents were simply asked in wave 9 to say in general whether their health is poor, fair, good, very good or excellent, with this question also asked in wave 14 (BHPS variable: HLSF1). Ordinal measures of self-assessed health status have been widely used in the health economics literature to explore the relationship between health and income (see, e.g. O'Donnell et al., 2015). To make the interpretation of results more intuitive, we reverse the numerical coding of the BHPS variables so that higher scores correspond to better health.

The general health question asked in waves 9 and 14 is the first item in the Short Form (SF) health survey, with version 1 of the 36 item questionnaire administered in both waves (SF-36: BHPS variables HLSF1-HLSF10D). The SF health survey is designed to measure functional health and well-being from the individual's point of view and is widely used in clinical trials (see Ware, 1993). The SF-6D preference-based algorithm (Brazier et al. 2002) was used to estimate health-related quality of life (HRQoL) from the responses to SF-36. The resultant cardinal measure is bounded in the unit interval with full health corresponding to a value of one and death implicitly assigned a value of zero.

3.3 Empirical results

Table 1 presents the main results from the analysis of headcount IRHS by NUTS 1 region for the two ordinal measures of general health, HLSTAT and HLSF1. Figure 1 plots the estimates of the normalised headcount index $\widehat{S}$, together with the associated 95% confidence intervals. The HLSTAT results suggest a slight upward trend in normalised headcount IRHS, rising from about 0.035 in the early 1990's to roughly 0.045 by 2008, while the two HLSF1 estimates for 1999 and 2004 both lie at the top end of this range. These positive estimates imply that population health in richer regions was statistically preferable on average to that in poorer regions: the healthier of any randomly chosen pair was more likely to be from the richer than the poorer region, conditional on the two individuals being from different regions and controlling for any demographic differences between the populations of the two regions. Alternatively, $\widehat{S}$ may be interpreted as the population-weighted average value of the rank-biserial correlation (Cureton, 1956) between health outcomes and economic prosperity across mutually distinct pairs of regions, which although small is nevertheless statistically significant in all years.

Further insight into the source of this IRHS can be gained from Table 2, which presents detailed results for HLSTAT in 2004 that may be taken as being typical of those obtained for both general health measures and all years. Regions are ordered from the poorest to the richest, with a coefficient of variation of regional mean incomes of 0.127. The values in the main body of the table are the pairwise identification indices: for example, the {NE, GL} entry of 0.120 implies that if two individuals had been randomly chosen from the standardised populations of the two regions then there was a 12% difference in the chances that the healthier of the pair would have been from Greater London rather than the North East, with the Londoner healthier in 40.6% of such comparisons, the North Easterner in 28.6% and the pair being equally healthy in the remaining 30.7% of matches. Indeed, health on this yardstick was substantially worse in the North East, the poorest region in Great Britain, than in all other regions − as indicated by the string of large and significantly positive values in the {NE} row. As a result, the contribution of the North East to the overall headcount index value of 0.0345 was 15.6% despite the population share of the region being only 4.2%. Conversely self-reported health in London, the richest region but containing some of the poorest districts in the country, was significantly worse than in a number of the other more prosperous British regions, leading to the (insignificantly) negative net contribution of London to the headcount index *S*.

One possible cause of the observed trend in headcount IRHS is changes in the demographic structure of the British population over the study period. To investigate this possibility, the indices were re-estimated with the cross-sectional weights for all years standardised on the basis of the population structure in Great Britain in 1991. These 'fixed population structure' estimates are reported in the final pair of columns in Table 1 and suggest that headcount IRHS would have been virtually constant between 1991 and 2008 if it had not been for changes in the composition of the British population by sex and age class.

Table 3 presents results from the analysis of income-related HRQoL stratification by NUTS 1 region for the two years in which the SF Health Survey was administered as part of the BHPS. The main estimates of the headcount indices reported in Panel A are all positive but appreciably smaller than the corresponding estimates in Table 1 for the general health measure HLSF1, which is also obtained from the SF health survey but is not used in the computation of the HRQoL measure. More specifically, the normalised index values imply that the difference in the chances that a representative resident from a richer region having a higher rather than a lower level of HRQoL than one from a poorer region was only of the order of $2.5\% - 3\%$ in comparison to the difference of about 4.5% for HLSF1.

One possible cause of this discrepancy between the HRQoL and HLSF1 estimates is that the latter is a discrete variable that can only take five possible values whereas the former may be considered to be a continuous variable as it can take up to several hundred discrete values in the range between 0.301 (the worst possible score for respondents) and 1. To explore the possible effect of discretization on the results, the indices were re-estimated with the HRQoL data recoded into five classes, where the class boundaries were chosen in the two waves such that the proportion of the British population falling into each class was the same as for the HLSF1 variable. The discretized estimates reported in Panel B are only marginally different from the main estimates, which may be taken to imply that the considerable difference between the HRQoL and HLSF1 results is not due to the categorical nature of the latter variable but rather reflects substantive differences in the constructs underlying the two measures of health status. Additional sensitivity tests (results not reported) show that the headcount index is generally robust to the grouping of the HRQoL data into quantiles so long as there are no fewer than about 5 health utility classes.

The regional pattern of identification for HRQoL (results not reported) was broadly similar to that shown for HLSTAT in Table 3. In particular, the HRQoL of a randomly

chosen North Easterner was more likely to have been significantly lower rather than higher than that of randomly chosen residents of virtually every other British region. More generally, the results imply that randomly chosen residents of richer regions would have been more likely to have higher rather than lower HRQoL than those of poorer regions, with all but two of the significant pairwise indices being positive. Furthermore, more prosperous regions were typically found to have had higher levels of HRQoL on average, though the rankings of regions by mean income and by mean HRQoL were not identical with a Spearman's rank correlation coefficient of 0.618 for 1999 and 0.909 for 2004. As a result, the 'pure' HRQoL stratification index values reported in Panel C are somewhat larger than the corresponding IRHS values, with roughly 75% of the total stratification in HRQoL between regions systematically associated with regional disparities in incomes.

Finally, Table 4 reports between-region HRQoL slope inequality index estimates, which reveal that the expected scale of regional health utility disparities between the poorest and richest regions was small compared to the national average of just over 0.8 in both years. The table also presents estimates of the population-weighted average pairwise effect size between mutually distinct regions, where the effect size is calculated for each pair as the mean HRQoL difference standardised by the corresponding pooled estimate of the within-region standard deviation in individual HRQoL. If health utility in all regions was normally distributed with common variance then these estimates would imply $\widehat{S}$ values of 0.0387 in 1999 and of 0.0529 in 2004 (see McGraw and Wong, 1992), which are reasonably close to the direct estimates if upwardly biased due to the violation of the normality assumption.

4. Discussion

The IRHS headcount index may be used to quantify differences in population health outcomes between the regions of a country, where the socioeconomic dimension is taken into

account by ranking the regions in terms of economic prosperity rather than health status. The index provides a measure that should be easy to explain to policymakers, being equal to the population-weighted mean difference in the probabilities that that the healthier of any two randomly chosen individuals will be from the richer rather than the poorer region from which they are drawn. Moreover it is also possible to estimate the contribution of individual regions to headcount IRHS with the further potential to identify the characteristics or factors that contribute to stratification.

The index depends on the degree to which the populations of different regions occupy well-defined strata in the national distribution of the health outcome. It thereby takes into consideration the degree of variation in health outcomes within as well as between regions, unlike conventional methods for the measurement of between-region health inequality such as the socioeconomic gradient. Nevertheless the methodology does not also take account of income variation within regions, with this remaining a topic for future research given evidence that health outcomes in the poorer regions of Britain are not only worse on average but also across the entire income distribution. For example, Marmot at al. (2010, Figure 2.9) shows that if one compares neighbourhoods with the same level of income deprivation then disability-free life expectancy is lower in the North East than in London at all levels of neighbourhood income deprivation.

The index is well-defined even if only ordinal data are available, which is often the case with survey measures of self-reported health, subjective well-being and life satisfaction, being directly applicable to polytomous categorical variables without the need for either dichotomisation or cardinalisation. If cardinal health data are available then the proposed approach to the measurement of IRHS might be extended in the manner of Allanson (2016) to take into account not only the incidence but also the depth and severity of stratification.

The measurement framework could also be used to analyse differences between population groups classified on the basis of class, gender or race rather than region.

The index is used to examine the evolution of income-related health differences between the regions of Great Britain between 1991 and 2008, where it should be noted that the results are sensitive to the chosen level of spatial aggregation. In particular, aggregation over regions with widely differing levels of average income relative to the national average will tend to result in lower levels of IRHS. For example, a country-level analysis of IRHS between England, Wales and Scotland (results not reported) yielded insignificant estimates of headcount IRHS in virtually all years. Conversely, an analysis at the local district level would reveal localised pockets of both income deprivation and health disadvantage within regions, which are partially masked in the current study based on regional average incomes.

The empirical findings reveal three main points of interest. First there is a significant positive association between regional health and income outcomes, with the resident of the richer region likely to be the healthier of any randomly chosen pair from two different regions. In particular, the North East stands out as having been both the poorest and least healthy region in Great Britain throughout the study period: for example the region accounted in 2004 for as much as 15.6% of HLSTAT and 13.1% of HRQoL headcount IRHS despite a population share of only 4.2%. Health outcomes were also significantly worse in Wales, Yorkshire & Humberside and East Midlands than in many of the more prosperous regions of Southern England, broadly supporting the notion of a North-South divide within England (cf. Whitehead 2014). Nevertheless, the low degree of pairwise identification due to the overlapping of regional health distributions will have had the effect of obscuring these systematic differences in population health between richer and poorer regions, which might otherwise have been the object of greater public and policy concern.

Second, the geographical pattern of variation in general health does not exactly fit the familiar map of regional disparities in life expectancy across Britain (see Office for National Statistics, 2015; National Records of Scotland, 2015). In particular, levels of general health in Scotland were indistinguishable from similarly prosperous regions in the rest of Britain despite Scotland having had the lowest life expectancy of any region in Britain over the entire study period, mirroring similar findings in Taulbut et al. (2013) for West Central Scotland. However, this individual result should not be taken to imply that regional levels of prosperity were more strongly correlated with health than with life expectancy. For example, the correlation of regional average income with average HRQoL and life expectancy was 0.493 and 0.574 respectively for men in 2004. Further work is required to understand the associations between morbidity, mortality and socioeconomic conditions at a regional level.

Third, the lack of any apparent trend in headcount IRHS after controlling for demographic changes points to the persistence of the root causes of the observed differences in general health between regions. Whitehead (2014, p.5) observes that these causes are the same across the country, resulting from differences between socioeconomic groups not only in terms of poverty but also in the power and resources needed for health, in exposure to health damaging environments and in opportunities to enjoy positive health factors and protective conditions. Additionally, population health in certain areas, most notably Northern England, Wales and Scotland, may have continued to have been affected by the legacy of heavy industry and its decline.

The empirical study could be extended using health data from Understanding Society, the successor study to the BHPS, though differences in the health questions between the two studies would limit comparability considerably. In particular, Understanding Society does not include the variable HLSTAT and only contains version 2 of the 12 item SF health survey, potentially limiting interest to HLSF1. The adoption of a longitudinal study design would

further allow for the inclusion of death as a separate health outcome category, providing the basis for an analysis of income-related stratification in healthy life expectancy. It would also be of interest to examine regional differences in other life outcomes, such as subjective well-being, life satisfaction or educational attainment. More generally, the measurement framework could be implemented with sub-regional (e.g. super output area) statistics used in place of individual data in the construction of the regional health distributions.

Acknowledgements

The author bears sole responsibility for the further analysis and interpretation of the British Household Panel Survey data employed in this study.

References

Allanson, P. (2016). Income stratification and the measurement of interdistributional inequality between multiple groups. *Review of Income and Wealth, forthcoming*.

Allison, R. A., & Forster, J. E. (2004). Measuring health inequality using qualitative data, *Journal of Health Economics,* 23, 505–524.

Bleichrodt, H., & Van Doorslaer, E. (2006). A welfare economics foundation for health inequality measurement. *Journal of Health Economics, 25*, 945–957.

Brazier, J., Roberts, J., & Deverill, M. (2002). The estimation of a preference-based measure of health from the SF-36. *Journal of Health Economics, 21*, 271–292.

Buck, N., Burton, J., Laurie, H., Lynn, P., & Noah Uhrig, S. C. (2006). *Quality Profile: British Household Panel Survey* (Version 2.0:Waves 1 to 13: 1991-2003). Retrieved from https://www.iser.essex.ac.uk/files/bhps/quality-profiles/BHPS-QP-01-03-06-v2.pdf.

Cureton, E. E. (1956). Rank-biserial correlation. *Psychometrika, 21*, 287-290.

Dagum, C. (1997). A New Approach to the Decomposition of the Gini Income Inequality Ratio. *Empirical Economics, 22*, 515-531.

De Baets, B., & De Meyer, H. (2007). Toward Graded and Nongraded Variants of Stochastic Dominance. In I. Batyrshin, L. Sheremetov and L. A. Zadeh (Eds.), *Perception-based Data Mining and Decision Making in Economics and Finance, Studies in Computational Intelligence* (Volume 36, pp.261–274). Berlin Heidelberg: Springer-Verlag.

De Baets, B., De Meyer, H., & De Loof, K. (2010). On the cycle-transitivity of the mutual rank probability relation of a poset. *Fuzzy Sets and Systems, 161*, 2695–2708.

De Schuymer, B., De Meyer, H., & De Baets, B. (2003). A fuzzy approach to stochastic dominance of random variables. In: T. Bildi, B. De Baets and O. Kaynak (Eds.), *Lecture Notes in Artificial Intelligence* (Volume 2715, pp.253–260). Berlin Heidelberg: Springer-Verlag.

Department for Work and Pensions. (2015). *Households Below Average Income: An analysis of the income distribution 1994/95 – 2013/14*. Retrieved from https://www.gov.uk/government/statistics/households-below-average-income-19941995-to-20132014.

Erreygers, G. (2009). Correcting the Concentration Index. *Journal of Health Economics, 28*, 504–515.

Jiménez-Rubio, D., Smith, P. C., & Van Doorslaer, E. (2006). Equity in health and health care in a decentralised context: Evidence from Canada. *Health Economics, 17*, 377-92.

Kobus, M. (2015). Polarization measurement for ordinal data. *Journal of Economic Inequality, 13*, 275-97.

Lasswell, T. E. (1965). *Class and stratum*. Boston: Houghton Mifflin.

Lv, G., Wang, Y., & Xu, Y. (2015). On a new class of measures for health inequality based on ordinal data. *Journal of Economic Inequality, 13*, 465-77.

Marmot M., Allen J., Goldblatt P., Boyce T., McNeish D., Grady M., & Geddes I. (2010). *Fair society, healthy lives*: The Marmot review. Retrieved from http://www.instituteofhealthequity.org/projects/fair-society-healthy-lives-the-marmot-review.

Marmot, M. (2013) *Health inequalities in the EU: Final report of a consortium*. Retrieved from http://ec.europa.eu/health/social_determinants/docs/healthinequalitiesineu_2013_en.pdf.

McGraw, K. O., & Wong, S. P. (1992). A common language effect size statistic. *Psychological Bulletin, III*, 361-365.

Montes, I., Martinetti, D., Diaz, S., & Montes, S. (2015). Interpretation of Statistical Preference in Terms of Location Parameters. *INFOR: Information Systems and Operational Research, 53*, 1-12.

National Records of Scotland. (2015). *Expectation of life, by sex and selected age, Scotland, 1861 to 2014*. Retrieved from www.gro-scotland.gov.uk/statistics-and-data/statistics/statistics-by-theme/life-expectancy/life-expectancy-at-scotland-level/scottish-interim-life-tables.

O'Donnell, O., Van Doorslaer, E., & Van Ourti, T. (2015). Health and Inequality. In A. B. Atkinson and F. J. Bourguignon (Eds.), *Handbook of Income Distribution, Volume 2B* (pp. 1419-1533). Amsterdam: Elsevier.

Office for National Statistics. (2015). *Life expectancy at birth and at age 65 by local areas in England and Wales, 1991-1993 to 2012-2014*. Retrieved from www.ons.gov.uk/ons/rel/subnational-health4/life-expectancy-at-birth-and-at-age-65-by-local-areas-in-england-and-wales/2012-14/index.html.

Office for National Statistics. (2016). *A Guide to Sources of Data on Earnings and Income*. Retrieved from https://www.ons.gov.uk/employmentandlabourmarket/peopleinwork/earningsandworkinghours/methodologies/aguidetosourcesofdataonearningsandincome.

Office for National Statistics. (various years). *Regional Trends.* London: The Stationery Office.

Taulbut M., Walsh D., Parcell S., Hartmann A., Poirier G., Strniskova D., Daniels G., & Hanlon P. (2013). What can ecological data tell us about reasons for divergence in health status between West Central Scotland and other regions of post-industrial Europe? *Public Health, 127*, 153-163.

University of Essex, Institute for Social and Economic Research. (2012). *British Household Panel Survey: Waves 1-18, 1991-2009* [computer file]. 7th Edition. Colchester, Essex: UK Data Archive [distributor], July. SN: 5151.

Ware, J. E., Snow, K. K., Kolinski, M., & Gandeck, B. (1993). *SF-36 Health survey manual and interpretation guide*. Boston, MA: The Health Institute, New England Medical Centre.

Wagstaff, A. (2005). The bounds of the concentration index when the variable of interest is binary, with an application to immunization inequality. *Health Economics, 14*, 429–432.

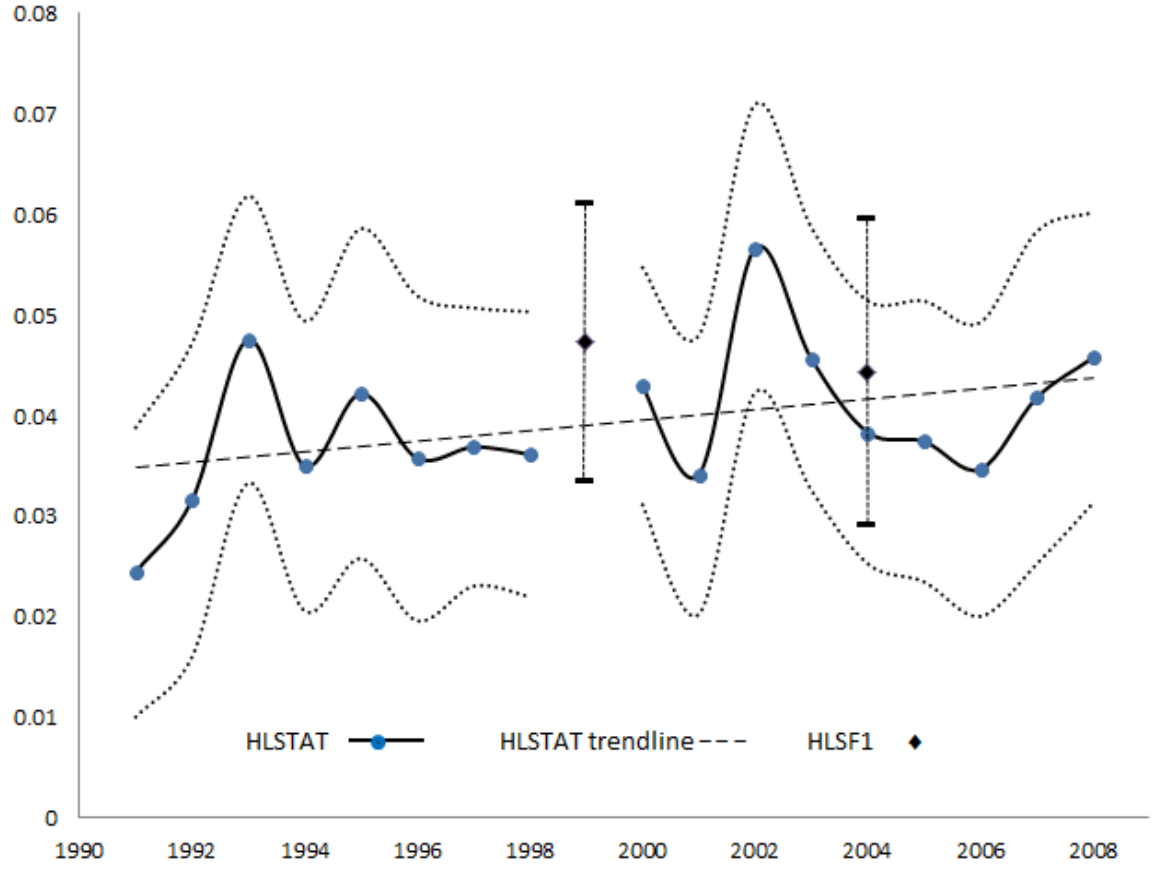
Wang, H., & Yu, Y. (2016). Increasing health inequality in China: An empirical study with ordinal data. *Journal of Economic Inequality, 14*, 41-61.

Whitehead M. (2014). *Due North: Report of the Inquiry on Health Equity for the North*. Retrieved from http://www.instituteofhealthequity.org/projects/due-north-the-report-of-the-inquiry-on-health-equity-for-the-north.

Yitzhaki, S., & Lerman, R. (1991). Income stratification and income inequality. *Review of Income and Wealth,* 37, 313-329.

Zheng, B. (2011). A new approach to measure socioeconomic inequality in health. *Journal of Economic Inequality, 9*, 555-577.

Figure 1. General health normalised headcount indices $\widehat{S}$ by NUTS 1 region, 1991-2008



Source: Own calculations. 95% confidence intervals constructed using bootstrapped standard errors based on 500 replications. Linear trendline fitted by least squares.

Table 1. General health IRHS headcount indices, 1991-2008

Year	Headcount index S		Normalised index $\widehat{S}$		$\widehat{S}$ based on 1991 population structure	
	HLSTAT	*HLSF1*	*HLSTAT*	*HLSF1*	*HLSTAT*	*HLSF1*
1991	0.0219		0.0244		0.0244	
	0.0066		*0.0074*		*0.0073*	
1992	0.0284		0.0315		0.0350	
	0.0072		*0.0080*		*0.0081*	
1993	0.0428		0.0475		0.0514	
	0.0066		*0.0073*		*0.0080*	
1994	0.0314		0.0349		0.0407	
	0.0066		*0.0074*		*0.0078*	
1995	0.0379		0.0421		0.0524	
	0.0075		*0.0084*		*0.0100*	
1996	0.0321		0.0357		0.0451	
	0.0074		*0.0083*		*0.0089*	
1997	0.0331		0.0368		0.0423	
	0.0064		*0.0071*		*0.0084*	
1998	0.0325		0.0361		0.0419	
	0.0065		*0.0072*		*0.0087*	
1999	-	0.0426	-	0.0473	-	0.0473
		0.0064		*0.0070*		*0.0070*
2000	0.0386		0.0429		0.0423	
	0.0054		*0.0060*		*0.0068*	
2001	0.0307		0.0340		0.0327	
	0.0064		*0.0071*		*0.0078*	
2002	0.0510		0.0566		0.0562	
	0.0066		*0.0073*		*0.0073*	
2003	0.0411		0.0456		0.0382	
	0.0060		*0.0067*		*0.0071*	
2004	0.0345	0.0399	0.0383	0.0443	0.0351	0.0409
	0.0060	*0.0070*	*0.0067*	*0.0078*	*0.0070*	*0.0081*
2005	0.0337		0.0374		0.0389	
	0.0064		*0.0071*		*0.0080*	
2006	0.0311		0.0346		0.0339	
	0.0067		*0.0075*		*0.0090*	
2007	0.0375		0.0417		0.0408	
	0.0076		*0.0084*		*0.0090*	
2008	0.0411		0.0457		0.0412	
	0.0066		*0.0074*		*0.0085*	

Source: Own calculations. Bootstrapped standard errors in italics based on 500 replications. All estimates are statistically significant at the 1% level..

Table 2: Detailed breakdown of the HLSTAT IRHS headcount index by NUTS 1 Region in 2004

Region		Popn share %	Mean equiv. income £/week	Pairwise identification indices											Regional index S_{row}	Share of S %
				NE	WA	WM	YH	NW	EM	SC	SW	EE	SE	GL		
North East	NE	4.2	478	0	0.104**	0.088*	0.081*	0.133**	0.102**	0.157**	0.142**	0.182**	0.174**	0.120**	0.0054**	15.6
					0.029	*0.035*	*0.040*	*0.029*	*0.034*	*0.029*	*0.030*	*0.035*	*0.030*	*0.034*	*0.0011*	
Wales	WA	5.2	485		0	-0.021	-0.023	0.027	-0.005	0.052**	0.038*	0.071**	0.063**	0.013	0.0015**	4.4
						0.023	*0.027*	*0.019*	*0.020*	*0.017*	*0.019*	*0.021*	*0.018*	*0.019*	*0.0006*	
West Midlands	WM	8.3	502			0	-0.003	0.050*	0.016	0.076**	0.061*	0.096**	0.089**	0.035	0.0040**	11.7
							0.032	*0.025*	*0.026*	*0.024*	*0.026*	*0.029*	*0.024*	*0.026*	*0.0013*	
Yorks & Humber	YH	9.0	502				0	0.051	0.019	0.076**	0.062*	0.097**	0.089**	0.037	0.0044**	12.8
								0.030	*0.030*	*0.029*	*0.029*	*0.032*	*0.030*	*0.030*	*0.0014*	
North West	NW	12.2	522					0	-0.033	0.025	0.012	0.044	0.036	-0.015	0.0030**	8.8
									0.023	*0.020*	*0.021*	*0.024*	*0.021*	*0.022*	*0.0009*	
East Midlands	EM	8.4	532						0	0.059**	0.045*	0.079**	0.071**	0.019	0.0028**	8.1
										0.021	*0.021*	*0.027*	*0.023*	*0.023*	*0.0005*	
Scotland	SC	8.9	536							0	-0.013	0.017	0.009	-0.040*	0.0026**	7.4
											0.020	*0.023*	*0.020*	*0.020*	*0.0005*	
South West	SW	9.3	550								0	0.031	0.023	-0.026	0.0025**	7.3
												0.024	*0.022*	*0.023*	*0.0005*	
East of England	EE	10.3	603									0	-0.008	-0.059*	0.0039**	11.3
													0.024	*0.026*	*0.0010*	
South East	SE	15.0	666										0	-0.051*	0.0051**	14.7
														0.021	*0.0016*	
London	GL	9.1	703											0	-0.0007	-2.1
															0.0014	
Headcount Index S															0.0345**	100
															0.0060	

*Source: Own calculations. Regions are ranked in order of mean weekly equivalised household income (before housing costs). The matrix is symmetric about the leading diagonal. Bootstrapped standard errors in italics based on 500 replications. Statistical significance at 1% and 5% levels are denoted by ** and * respectively.*

Table 3. HRQoL IRHS headcount indices, 1999 and 2004

	Headcount index S	Normalised index $\widehat{S}$
A: Main IRHS results		
1999	0.0222	0.0247
	0.0055	*0.0061*
2004	0.0275	0.0305
	0.0056	*0.0062*
B: IRHS with discretized HRQoL data		
1999	0.0239	0.0265
	0.0065	*0.0072*
2004	0.0280	0.0310
	0.0055	*0.0060*
C: 'Pure' HRQoL stratification		
1999	0.0306	0.0340
	0.0057	*0.0064*
2004	0.0354	0.0394
	0.0068	*0.0076*

Source: Own calculations. See Table 1 for notes.

Table 4. HRQoL cardinal indices, 1999 and 2004

	1999	2004
Between-region Slope Inequality Index	0.0185	0.0246
	0.0042	*0.0047*
Population-weighted average pairwise effect size	0.0486	0.0663
(over distinct pairs of regions)	*0.0112*	*0.0129*

Source: Own calculations. See Table 1 for notes.